A New Method to Convert the DNA Sequence of Human to a QR Code

Sadoon Hussein Abdullah

saadustego@gmail.com

Department of Science Biology,
Collage of Science, Mosul University

Zahraa Tariq Abdulhamid

zahtarq@gmail.com

Department of Science Biology, Collage of
Education for Girls, Mosul University

ABSTRACT

Data transmission, security and authenticity, nowadays is considered the major source of problems. To face these problems a suitable solution have to be forwarded . Steganography is forwarded one of the solutions as a technique among many. This paper includes proposal for using a new technique called (Quick Respond Codes) QR Codes .

QR Codes technique is mainly used to convert and store messages since it has higher and large storage capacity, in addition to that we used, in this paper , the STR (Short Tandem Repeats) DNA sequence. STR DNA is very important because each person has its own DNA which means that security and authentication (reliable) are very high. The proposed method is to convert the hidden data (STR DNA sequence) to QR code that mean the security is at maximum. Then, by using the MD5 (Message Digest 5) Algorithm, the result data is converted to QR code and the data within the data areas are stored in an unusually large format to make retrieval difficult.

Finally, STR DNA sequences were applied in different lengths and converted to QR code. By this method , the QR code is applied in different applications and it was very fast (3-4 second). To apply this technique and to reduce the cost, we use a webcam to get back the private (confidential) data from the QR code. The data were retrieved completely without any error because QR code is reliable and authentic. From the above result we suggest to use QR Code for indentification of the of the individual (ID).

Keywords: DNA, Quick Response code, Encrypt.

تحويل تتابعات الحامض النووي للإنسان إلى رمز QR بأسلوب جديدة

زهراء طارق عبد الحميد

قسم علوم الحاسوب

كلية التربية للبنات

جامعة الموصل، الموصل، العراق

سعدون حسين عبد الله

قسم علوم الحياة

كلية العلوم

جامعة الموصل، الموصل، العراق

الملخص

ان نقل البيانـات مـن الناحيـة الامنيـة والوثوقيـة (التخويـل) تعـد فـي ايامنـا هـذه مـن المشـاكل الرئيسـية ولمواجهـة هـذه المشـاكل يجـب تقـديم الحلـول المناسـبة، ان تقنيـة علـم اخفـاء البيانـات (Steganography)

هو واحد من التقنيات العديدة المستخدمة في هذا المجال. تتضمن هذا البحث اقتراح طريقة لاستخدام تقنية جديدة تعرف برمز الاستجابة السريعة (QR Code).

ان رمز الاستجابة السريعة يستخدم في الدرجة الرئيسية كغطاء وخزن البيانات (الرسائل) لان يتمتع بقابلية كبيرة على خزن البيانات، وفضلا عن ذلك فقد استخدمنا في هذا البحث تتابع الحامض النووي STR ، يعد هذا التتابع مهم جدا لان كل شخص له تتابع الحامض النووي STR خاص به وبذلك ان السرية والوثوقية (التخويل) تكون عالية جدا. ان الطريقة المقترحة تحول البيانات المخفية (تتابع الحامض النووي STR) الى رمز الاستجابة السريعة QR وذلك يعني ان السرية تكون في اعظم حالتها. وباستخدام خوارزمية MD5 فان البيانات (تتابعات دنا STR) يتحول الى رمز QR والبيانات تخزن ضمن رمز QR بشكل غير مالوف (غير قياسي) لجعل استرجاع البيانات صعبة.

واخيرا، خلال هذا البحث تم تطبيق تتابعات دنا STR و باطوال مختلفة ثم تم تحويله الى رمز QR وبهذه الطريقة فان رمز QR جرى تنفيده في تطبيقات مختلفة وكان عملية الترميز وعملية الاسترجاع سريعة جدا (3-4) ثواني بالترتيب، ولتطبيق هذه التقنية ولتقليل الكلفة تم استخدام كامرة الحاسوب(Webcam) لاسترجاع البيانات السرية من رمز QR، ولقد تم استرجاع جميع البيانات بدون اي اخطاء، من هذه النتائج المذكوره اعلاه نقترح استخدام رمز QR في بطاقة الشخصية (ID).

الكلمات المفتاحية: تتابعات الحامض النووي، رمز الاستجابة السريعة، تشفير.

1. Introduction

A barcode is an optic-image code which consists of areas with black and white parallel lines. These lines are an example of related data to a person or an object and by varying widths and spacing each person will have his own information as a barcode which will be readable by the camera, a smartphone or machine[1]. Fig.1 indicates the barcode

Fig.1.Barcode

Types	QR Code	Barcode

QR code which refers to Quick Response Code was feigned by the Japanese company Denso Wave as a trademark. This code has many features as [2]:

1. large data encoding,
2. resistance to damage
3. speed reading
4. small size image
5. 360 degrees reading

The table (1) indicates Comparison of Barcode and QR Code [2]

Density of Information	Higher	Lower
Capacity of Information	Biggest	Smallest
Information Type	Characters ,Number, , Chinese, Greek , Pictures, voice and other binary information	Number , Characters, Greek
Based on database	Does not base on database or telecommunication network	Must base on database or telecommunication network

Table1. Comparison of Barcode and QR Code

Firstly, these codes are extremely popular in Asian countries like Japan and becoming more and more in western countries because of their huge capacity for encoding data horizontally and vertically. The following tables show the characteristics of both kinds. The table2 is related to the QR Code (version 40) which shows the maximum number of characters encoded in and minimum error correcting level [3],The table 2. shows the capacity of QR codes. See Fig.2 that illustration of QR Code Structure [4].

Table 2. The capacity of QR codes

S. No.	Data Type	Characters
1	Number data	7,089
2	Characters data	4,296
3	8-bit byte data	2,953
4	Japanese and Chinese characters data	1,817

All QR version:
1. A distinguished black and white points for composition of the QR code.
2. Versions domain from Version1 (21* 21 points) to Version 40 (177* 177 points).
3. The QR code Version information can be specified in the part of a QR code followed in blue color.
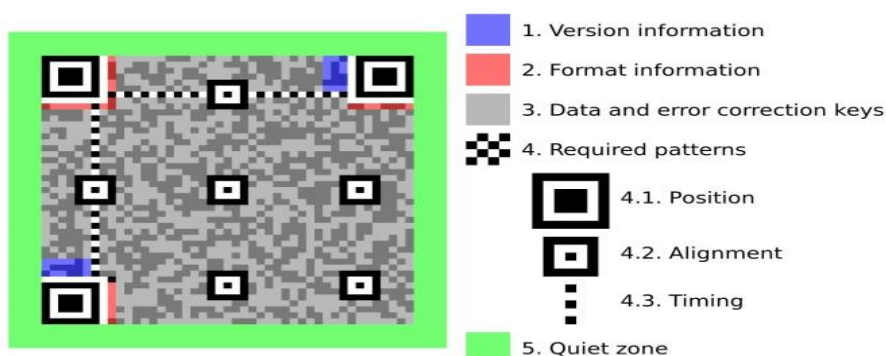

Fig.2. Structural QR Code

There are many characteristics of QR code [5]:

1. High Storage Capacity
   QR code storage is up to 7089 symbol of information, which is a large amount as compared
   to 1-D Barcode.
2. Encodable Character Set
   - Alphanumeric Characters (upper case letters A-Z; Digits 0 - 9; nine other characters
   - Japanese and Chinese characters data
   - Numeric data (Digits 0-9)
   - space, : % * + - / _ $)
3. Small Printout Size
   This characteristic gives the QR Code the ability to store the same amount of data both
   horizontally and vertically and the space acquired is one-four time.
4. 360 Degree Reading
   Another provided feature of QR Code is that it is readable from any direction.

2. Overview
a. DNA Structure

DNA, or deoxyribonucleic acid is made-up of molecules called nucleotides. All nucleotide contains a sugar group ,a phosphate group, and nitrogen base.The information in DNA is stored as a code made up of four chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T).The sequence of these bases marked the information available for organism. Human DNA has about 3 billion bases, and more than 99% of those bases are the same in all people.

Similar to the way the order of letters in the alphabet can be used to form a word, Another type of nucleic acid, ribonucleic acid, or RNA, convert genetic information of DNA into proteins. So the sequence of bases can act as a code, providing the prescript for many biological functions.

In 1984, Alec Jeffreys first developed his groundbreaking technique using DNA variations to establish unique human identities, Nuclear DNA signs, such as short tandem repeats (STR), are widely applied for crime realization and parenthood testing [6].

b. MD5 algorithm

In 1991, Ron Rivest designed MD5 which stands for (Message Digest Algorithm 5) which means to process a variable length data into a fixed length message output of 128 bits.
MD5 has a mix of communal functions.
   - It 512 bits on blocks.
   - It processes each block through 4 rounds where each round, in turn, processes 16 sub blocks.
   - Before processing The 512 bit of message is divided into 16 sub blocks..
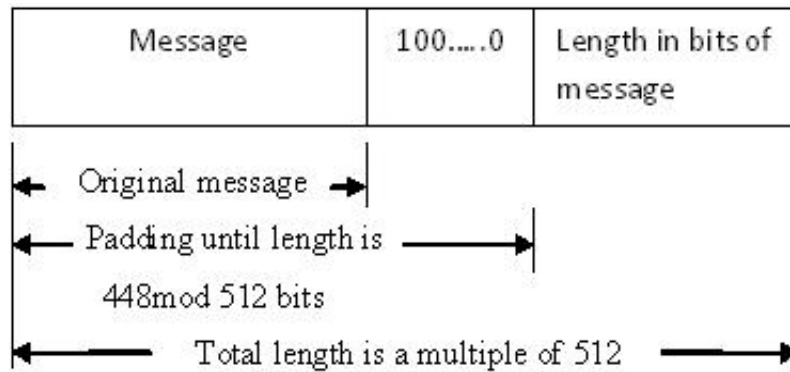Fig.3 [7] indicates the Length of the message after padding .

Fig. 3. Length of the message after padding (in bits)

3.Related Work

This section is a representation of an overview of related papers and techniques:

- In reference [8] they proposed a solution to the QR code recognize problem caused by normal camera device collection. They put forward the recognize algorithm and another recognition algorithm based on image processing.

- In reference [9] the algorithm which join corner detection with hull algorithm. Experimental realization made that the algorithm planned by the paper can correctly find the four apexes of the QR code and achieve good geometric correction. It will also increase the recognition rate of distorted QR code images.

- On the other hand, in reference [10]. They presented a method for 2D barcode image decoding in Identification and capture of data . In their paper, they survey the traditional decoding procedure by proposing a well designed image preprocessing method. The decoding procedure consists of image binarization, QR code capturing, perspective transformation and resampling process, and error. By these steps, they tried to recognize different types of QR code images.

- In reference [11]. They offered a solution for encoding/decoding DNA information in 2D barcodes.

- In reference [12] DNA barcoding makes it possible to recognize meat species of neatly related animals. Species identification of animal material in food samples is fundamental to authenticate the composition of meat products from illegal exchange.

4. Proposed Work

The process of converting the unique DNA sequence (STR) of humans goes through several steps until it get a QR code that uses in different applications which will be considered as a logo for anyone because it is unique.

C # languge was used in the proposed research, and the specifications of the computer that were worked on are:Core i3 processor, 4G RAM,1G internal graphics card and Operating System is Windows10.

Algorithm Details

1. Encoding STR DNA sequence to QR Code steps

The following steps are to convert STR sequence to QR code using different algorithms to increase security while getting a QR code in Jpeg format code to be used in different applications.

Step1: This step is done in biologic laboratories to obtain a unique DNA sequence and is placed in a specific database (STR Database). In this research, biological databases were used on the Internet to give samples of DNA sequences.

Step2: Convert STR sequence (S) to a string of (byte string) stream of binary (ST).

Step3: Apply MD5 on ST to get high security and confidentiality (MDST).

Steps4: The MDST is converted to a QR code depending on the proposed method, the data in MDST placed in QR data area (data zone) after data fragmentation in these places. The data will be hidden within the QR code according to specific contexts so that the data is distributed in a non-standard way , Hash process is based on storing the data sites to retrieve them in the retrieval phase of the original data to prove reliability (authorized), High level mode and different properties are used to organize the data in a QR code image (QRimg) ,The QRimg is in jpeg format (*.JPEG).

Step5: There are three possibilities either printed QRimg or inserted into an application (Here it was placed in ID) or stored in a specific database.

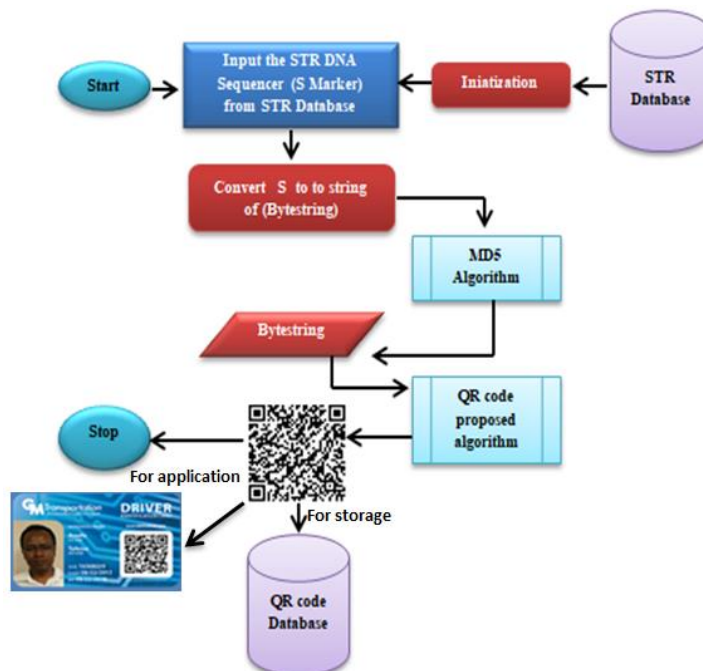The Fig.4. illustrates Block diagram of Encoding procedure , These steps are done by the supervisor(admin).

Fig.4. Block diagram of Encoding procedure

2.Decoding and authorized QR Code (Taken from personal identity ID) steps

To a scertain that the person who carries a QR code Owned depends on the DNA STR sequence for him and know its return reliability (authorized) applied the steps:

Step1: capture QR Code from ID card by using laptop webcam and save as JPEG
extension (CP.JPEG).
Step2: CP.JPEG is processed by propose algorithm to retrieve the contained data based
on the data storage location (QRData).
Step3: Insert QRData to MD5 algorithm and produce new bytestring of data
(MDdata).
Step4: Convert MDdata to STR DNA sequence format (STR).
Step5: To check the identity compare the STR with main database to authorize or not
authorized person.
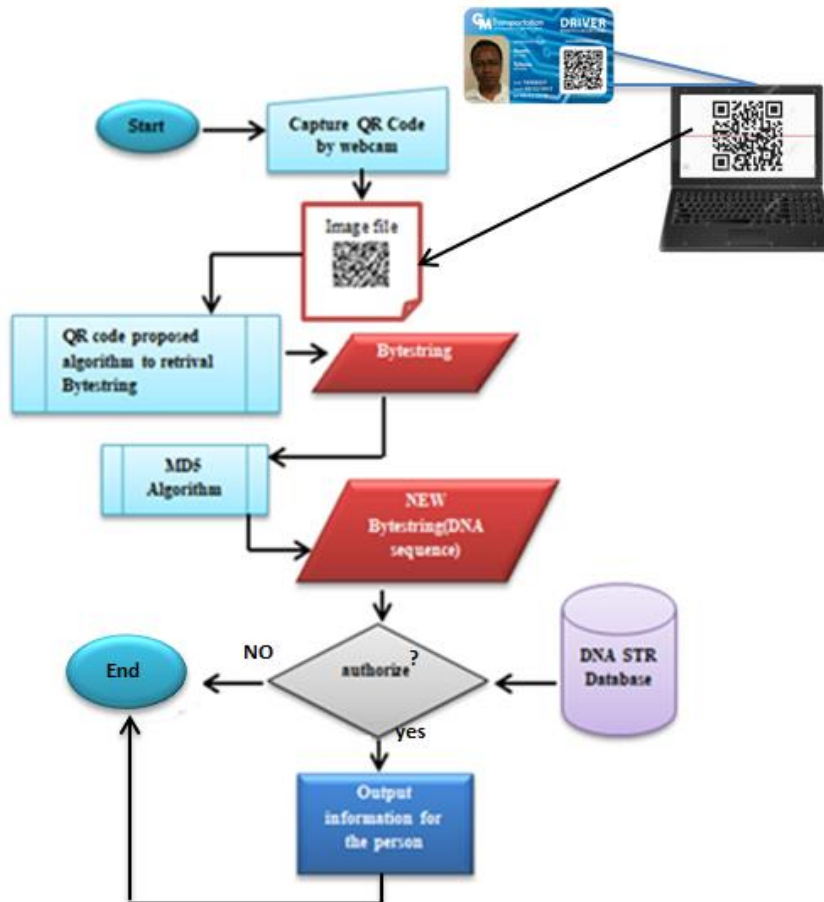The Fig.5. illustrates Block diagram of Decoding procedure .

Fig.5.Flow Chart of Decoding procedure

5. Results

The proposed method is better than the method in reference [11] since in
reference [11] our DNA sequence is converted into barcode either proposed method
the DNA sequences converted into QR code by high capacity data and high secrecy
while the possibility of scattered data into QR code.

While the proposed method the DNA sequence has been converted to QR code
and didn't abide by confidentiality issues for data and data is embedded directly within
the QR code either proposed method was the best from reference [12] where MD5
algorithm was added to increase secrecy and to difficulty retrieving secure data within
the QR code arranged data within QR code. Proposed method to comfort several
properties including: The table 3. shows the details of results with related methods.

- **Speed:** by applying different example to proposed method show it was faster than other algorithm ,so it appear that it save time ,the Encoding steps took very little time through all the steps where the time taken was (3) seconds while the Decoding steps took (4) seconds.

- **little Space** :by applying different length of STR DNA sequence to the proposed algorithm show included within QR code was greater than the barcode so it appear that is save space, and the size could be increased by using different types of QR code.

- **Authorize:** The proposed method applied them to different examples and gave high reliability due to using STR DNA person and also retrieve data within QR code and compared to the database to prove it**.**

- **Security :** The proposed method gave highly secure data:
  1. Because using STR unique made QR code also unique in quality and good character for each person.
  2. The use of the MD5 algorithms give strong secrecy to the DNA sequence in both cases Encoding and Decoding.
  3. The distribution of data within QR code and irregularly also gave high confidentiality to prevent the retrieval of secure data.

- **Different applications :**This proposed method would be appropriate in any business, and can use this method to keep his important documents, such as passport number, card ID. Etc.

- **Data recovery :**All data that is stored into QR code retrieved and 100% meaning that data
  is not lost during worked Encoding / Decoding.

- **Tools :** You can use some of the computer tools, for example a webcam instead of other
  devices to read the QR code for several reasons**:**
  1- In proposed method used Webcam an input of a QR code image directly to the program designed for the existence of codes are handled directly.
  2- The availability of a webcam within your computer provides the buy of private reading devices.
  3- QR code scanner read the standard data within a code, but if the data positions are changed within them, the final results are not understood or true.

Table 3. The details of results with related methods

| Item | Proposed method | | related method references [8,9,10,11,12] |
	Attribute	The Details	
Speed (Time)	High speed	Encoding (3 sec.),Decoding (4 sec.)	Did not mention
Security ability	Yes	uesd STR Sequence and MD5 algorthim	[11,12] used spacial DNA Sequences
Reduce cost tools	Yes	Use the webcam instead of other devices	Uesd normal Cameras (scanner device)
Length of Sequence (STR)	Different length	153 char's to 1455 char's	Did not mention
Bit Error Rate	0%	Retrieve all data	Did not mention
Authorize	Yes	Uesd uniqe STR Sequence	[12] only
The use (Applied)	Yes	Applied to the ID card	Different application
The Type of QR Code image	Different Type	JPEG (It has high resolution in the picture)	Did not mention
Type of Code	QR Code	It was used for the above reasons	[8,9] used QR Code,[10,11,12] used Barcode

6. Conclusions

- The research shows that a QR code contains a large area of information storage, and dealing with these areas is easy and can be applied to several applications.

- According to the results shown in the proposed research, the use of str DNA sequences gave high confidentiality and reliability to the resulting QR code because it is unique to each person and increased the secret of the resulting QR code by inserting DNA str sequences into the Md5 algorithm.

- In this study, the reliability of the QR code holder was proved by matching the sequences of str DNA restored with the database by using one of the alignment methods which showed that the data was completely restored and matched with the database. That is to say the value of BER = 0.

- In this research, the cost was reduced by using the laptop camera, instead of the scanner, which has the ability to read the data because it contains QR code that has the most modern programming languages. Thus any researcher can get benefit from this characteristic.

- The measured time from the beginning of the algorithm to obtain the QR Code, which is 4 seconds, is relatively long that depends on the length of the STR DNA sequences.While the time to restore the data and to prove the code and its reliability are relatively 3 seconds.

REFERENCES

[1]   Kinjal H. Pandya, ,Hiren J. Galiyawala, "A Survey on QR Codes: in context of Research and Application", Volume 4, Issue 3, March 2014,p258-262, India.

[2]   Md. Sanaul Haque, Richard Dybowski," Advanced QR Code Based Identity Card: A New Era for Generating Student ID Card in Developing Countries", First International Conference on Systems Informatics, Modelling and Simulation ,2014,p76-82,UK.

[3]   M. Mary Shanthi Rani,K.Rosemary Euphrasia,"Data Securitthrough QR Code Encryption And Steganography", Advanced Computing: An International Journal (ACIJ), Vol.7, No.1/2, March 2016,p1-7, India.

[4]   Julian Brackins,Mengyu Qiao,"A Secure QR Code Scheme",2017.

[5]   Sudeep S. Kulkarni and Chetna Malagi," Creation and Analysis of QR Code", Bonfring International Journal of Software Engineering and Soft Computing, Vol. 6, Special Issue, October 2016,p86-89.

[6]   Forensic DNA analysis: a primer for courts Issued: November 2017 DES4928 This primer is produced by the Royal Societyand the Royal Society of Edinburgh in conjunction with the Judicial College, the Judicial Institute, and the Judicial Studies Board for Northern Ireland.

[7]   Mary Cindy Ah Kioon, ZhaoShun Wang and Shubra Deb Das," Security Analysis of MD5 algorithm in Password Storage**,** Proceedings of the 2nd International Symposium on Computer, Communication, Control and Automation (ISCCCA-13),2013, China.

[8]   Gu, Yunhua, and Weixiang Zhang. "QR code recognition based on image processing." Information Science and Technology (ICIST), 2011 International Conference on. IEEE, 2011.

[9]   Kong, Suran. "QR code image correction based on corner detection and convex hull algorithm." Journal of Multimedia 8.6 ,2013, p662-668.

[10]   Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2013, Article ID 848276, 10 pages   http://dx.doi.org/10.1155/2013/848276.

[11]   Elena Purcaru, Cristian Toma," 2D Barcode for DNA Encoding", Journal of Mobile, Embedded and Distributed Systems, vol. III, no. 3, 2011,p142-153.

[12]   Tigor Naul," DNA QR Code Scanner for Identifying the Species Origin of Meat Products", Tigor Nauli/ Information Systems International Conference 2015, Indonesia.